MULTIMEDIA UNIVERSITY

STUDENT ID NO

MULTIMEDIA UNIVERSITY

FINAL EXAMINATION

TRIMESTER 2, 2022/2023

DSP5014 – STATISTICS AND PROBABILITY

(for diploma students only)

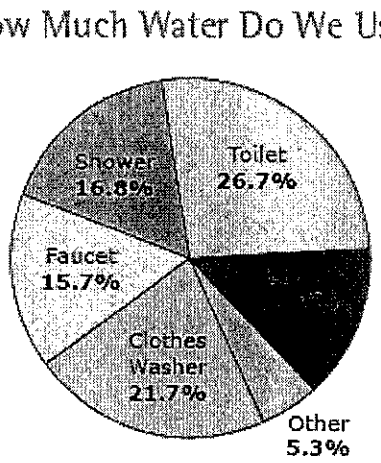
13 JULY 2023
9.00 a.m – 11.00 a.m
(2 Hours)

INSTRUCTIONS TO STUDENT

1. This question paper consists of 5 pages including appendix.
2. Attempt **ALL** questions. The distribution of the marks for each question is given.
3. Please write all your answer in the Answer Booklet provided.
4. Key formulae are given in the Appendix.

QUESTION 1

a) The following chart shows water usage for a household per month.



Based on the chart above,

i. Determine the type of data for activity that involves usage of water. (1 mark)

ii. What kind of activity uses water the most? (1 mark)

iii. How much are the expenses to cover the leaking if the household allocates RM150 for the water usage? (2 marks)

iv. In your opinion, how we can reduce the water usage for a household. (2 marks)

b) Given data set of Covid-19 daily cases in Melaka for May 2022.

Covid cases in Melaka	Frequency	Relative Frequency	Midpoint
$0 \leqslant 30$	3		
$30 \leqslant 60$	7		
$60 \leqslant 90$	6		
$90 \leqslant 120$	9		
$120 \leqslant 150$	5		
Total			

i. Complete the group data table above. (3 marks)

ii. From the table, calculate the value of mean and variance. (6 marks)

Continue...

c) The following data are the total quiz marks for 20 students in class Beta.

20	13	30	24	16
30	15	28	28	17
29	19	24	34	11
35	37	26	31	22

i. Construct a stem and leaf diagram, then identify the distribution shape. (4 marks)
ii. Find the Q_1, Q_2 and Q_3. (5 marks)
iii. Compute the lower and upper limit. (5 marks)
iv. Is there any potential outlier? (1 mark)

[TOTAL: 30 MARKS]

QUESTION 2

a) Suppose two fair dice are rolled. List out the possible outcomes for
i. the sum of dice is more than 5. (5 marks)
ii. the sum of dice is 8. (5 marks)

b) There are 3 production lines involved for the manufacturing processes. The table below provides the information on the process duration for each production line.

Line	Painting	Cutting	Packaging
A	13	2	20
B	16	4	10
C	14	6	15

Find the probability if a random product chosen
i. produces from line C. (2 marks)
ii. produces from line B and involves packaging process. (2 marks)
iii. involve cutting process or from line C. (3 marks)
iv. involves painting process given that it is produced from line A. (3 marks)

[TOTAL: 20 MARKS]

Continued ...

QUESTION 3

a) The mean length of the wood is supposed to be 6.5 feet. A builder wants to check whether the shipment of wood he receives has a mean length less than 6.5 feet. If the builder observes that the sample mean of 81 pieces of lumber is 6.3 feet with a population standard deviation of 1.7 feet. What will he conclude at 10% significance level? (12 marks)

b) The CEO of a large computer company claims that 80% of his customers are "very satisfied" with the customer service they receive. To test this claim, the researcher surveyed 150 customers and 117 of them stated they are "very satisfied." Based on these findings, can we reject the CEO's hypothesis that 80% of the customers are very satisfied? Use 5% significance level. (13 marks)

[TOTAL: 25 MARKS]

QUESTION 4

You want to examine the relationship between the age and price for used cars sold in the last year by a car dealership company. The results as following.

Car Ages (Years)	4	5	7	7	8	9	9	10	11	12
Price (RM)	6300	5800	4700	4500	4200	4000	3100	2100	2500	2200

a) Calculate SS_{xx} , SS_{yy} and SS_{xy} based on the information given. (9 marks)

b) Calculate the correlation coefficient, r and interpret the answer. (4 marks)

c) Find the linear regression equation $\hat{y} = \beta_0 + \beta_1 x$. (5 marks)

d) Predict the price if the used car ages are 6 years. (2 marks)

e) Compute the value of *SST*, *SSR* and *SSE*. (3 marks)

f) Calculate the standard error of the estimate, s_e . (2 marks)

[TOTAL: 25 MARKS]

End of Paper

APPENDIX – KEY FORMULA

	Ungrouped Data	Grouped Data
Mean	$\bar{x} = \frac{\sum x_i}{n}$ where n : sample size	$\bar{x} = \frac{\sum m_i f_i}{\sum f_i}$ where m : class midpoint, f : class frequency
Variance	$s^2 = \frac{1}{n-1}\left[\sum x_i^{\,2} - \frac{\left(\sum x_i\right)^2}{n}\right]$ where n : sample size	$s^2 = \frac{1}{\left(\sum f_i\right)-1}\left[\sum m_i^{\,2} f_i - \frac{\left(\sum m_i f_i\right)^2}{\sum f_i}\right]$ where m : class midpoint, f : class frequency

NORMAL AND STANDARD NORMAL PROBABILITY DISTRIBUTION

- z-value (observed value) for an x value : $Z = \frac{x-\mu}{\sigma}$

HYPOTHESIS TESTING

Population	Mean	Proportion
1	$Z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$	$Z = \frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$ where $\hat{p} = \frac{x}{n}$
2	$Z = \frac{\left(\bar{x}_1-\bar{x}_2\right)-\left(\mu_1-\mu_2\right)}{\sqrt{\frac{\sigma_1^{\,2}}{n_1}+\frac{\sigma_2^{\,2}}{n_2}}}$	$Z = \frac{\left(\hat{p}_1-\hat{p}_2\right)-\left(p_1-p_2\right)}{\sqrt{\hat{p}_p\left(1-\hat{p}_p\right)}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$ where $\hat{p}_p = \frac{x_1+x_2}{n_1+n_2}$

LINEAR REGRESSION ANALYSIS

- **Least square regression equation,** $\hat{y} = \beta_0 + \beta_1 x$ where $\beta_1 = \dfrac{S_{xy}}{S_{xx}}$ and $\beta_0 = \bar{y} - \beta_1 \bar{x}$

- **Correlation coefficient,** $r = \dfrac{S_{xy}}{\sqrt{S_{xx} S_{yy}}}$ where

$S_{xx} = \sum x^2 - \dfrac{(\sum x)^2}{n}$	$S_{yy} = \sum y^2 - \dfrac{(\sum y)^2}{n}$	$S_{xy} = \sum xy - \dfrac{(\sum x)(\sum y)}{n}$

- **Sum of Square**

$SST = SSR + SSE$	$SST = S_{yy}$	$SSR = \dfrac{(S_{xy})^2}{S_{xx}}$

- **Test of Significance for Regression Slope**

Standard Error, $s_e = \sqrt{\dfrac{SSE}{n-2}}$	Regression t-Test $t = \dfrac{\beta_1}{s_e / \sqrt{Sxx}}$	Correlation t-Test $t = r \cdot \sqrt{\dfrac{n-2}{1-r^2}}$